Breast Cancer Classification Model Performance Analysis

Models Comparison and Performance Evaluation

After training and evaluating the three different machine learning models (Logistic Regression, Random Forest, and Decision Tree developed to learn from the Scikit-Learn breast cancer dataset, It can be concluded that the Logistic Regression performs the best. Logistics Regression scores higher on the performance evaluation metrics and has a higher overall score outperforming the other two models.

Comparison Table

| Model | Accuracy | Precision | Recall | F1 Score | ROC - AUC | Overall Score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.9737 | 0.9722 | 0.9895 | 0.9790 | 0.9974 | 0.9824 |
| Random Forest | 0.9649 | 0.9589 | 0.9859 | 0.9722 | 0.9953 | 0.9754 |
| Decision Tree | 0.9474 | 0.9577 | 0.9577 | 0.9577 | 0.9440 | 0.9529 |

Module Evaluation

Logistic Regression

* Has the highest accuracy and ROC-AUC demonstrating its ability to correctly classify breast cancer tumors (benign and malignant tumors)
* Also has good precision and recall abilities demonstrating that the model is very capable of identifying true positives and false negatives

Random Forest

* This model also performs quite well but it is not able to match the performance of the Logistic Regression Model
* It has the second-best performance and provides reliable solutions

Decision Tree

* Does not perform as well as the other two models and has the weakest overall performance
* The model performance is limited and does not distinguish the between the data as well as the other two models decreasing the model's correctness

Overall the models did very well but the Logistic Regression demonstrates the best functionality with the highest success overall compared to the other two models. It consistently achieves higher scores across the evaluation metrics (comparison table), demonstrating its robustness in handling this type of binary classification problem.